# On a mathematical relation between the Eigen model and the asexual Wright-Fisher model

Fabio Musso, Departamento de Física, Universidad de Burgos

Facultad de Ciencias, Plaza Misael Bañuelos s/n, 09001 Burgos, Spain.

Tel: +34947258894; Fax: +34947258831.

E-mail address: fmusso@ubu.es

**Abstract**

We show that the Eigen model and the asexual Wright-Fisher model can be obtained as different limit cases of a unique stochastic model. This derivation makes clear which are the exact differences between these two models.

The two key concepts introduced with the Eigen model, the error threshold and the quasispecies, are not affected by these differences, so that they are naturally present also in population genetics models. According to this fact, in the last part of the paper, we use the classical diploid mutation-selection equation and the single peak fitness approximation to obtain the error threshold for sexual diploids. Finally, we compare the results with the asexual case.

**Keywords:** Mutation-selection dynamics, Error threshold, Stochastic model.

## 1 Introduction

The Eigen model was formulated as a deterministic mutation-selection model describing replication at the onset of life [4]. The study of mutation-selection balance in the Eigen model for very high mutation rates led to the development of two new evolutionary concepts: the error threshold and the quasispecies [5]. The first refers to the fact that, for a critical value of the mutation probability (and for some choices of the fitness landscape, see [2], [17], [15]), there is an abrupt transition in the asymptotic state of the system from a cloud of mutants organized around a given consensus sequence to an almost random distribution of genotypes. The second refers to the fact that, due to

the mutational coupling, selection acts on groups of neighbour mutants (called quasispecies) instead of individuals.

Since RNA viruses lack proof reading mechanisms, they have mutation rates orders of magnitude higher than DNA based organisms, so that both the error threshold and the quasispecies concepts could play a relevant role for these organisms. Indeed, the Eigen model has became the main mathematical tool in this context (see [10], [11] for recent reviews on the subject). However, some authors questioned the relevance of the quasispecies as a paradigm for populations of RNA viruses [12], [9], [7], [8], suggesting that the high heterogeneity in populations of RNA viruses could be due to genetic drift, and consequently could be better explained by population genetics models. This contrast could lead to the idea that the Eigen model and population genetic models are incompatible mathematical models. For example, [6] begins saying: "Some major differences distinguish quasispecies theory from the classical selection theories of Darwin and neo-Darwinian geneticists", while in [10], one can read: "The evolutionary dynamics of RNA viruses are complex and their high mutation rates, rapid replication kinetics, and large population sizes present a challenge to traditional population genetics". On the other hand, Wilke provided evidence that this is not the case, by showing that particular limit cases of the Eigen model give raise to some well known population genetics equations [17]. However, the precise mathematical relation between the Eigen model and population genetics models remains still unclear. The main purpose of the present paper is to fill this gap, by showing that the Eigen model and the haploid asexual Wright-Fisher model can be obtained as different particular limits of a unique discrete time stochastic model. This is done in section 2. Motivated by the analogies between the Eigen model and the Wright-Fisher model, in section 3 we use the classical diploid mutation-selection equation and the single peak fitness landscape approximation, to determine the error threshold for sexual diploids. Finally, to determine the influence of syngamy on the error threshold, in section 4 we compare the results that we obtained for sexual diploids with those holding for asexual diploids.

## 2   The stochastic model

In [13] it was shown that the Eigen model emerges as the deterministic and continuous time limit of a stochastic mutation-selection model. By changing the selection procedure of that model, we can obtain another stochastic model having again the Eigen model as its deterministic and continuous time limit and, at the same time, the asexual Wright-Fisher model as a different particular subcase.

Let us consider a population of constant size of $N$ individuals of $K$ possible different types $I_1, \ldots, I_K$ that reproduce asexually. Let $A_i$ be the fecundity of type $I_i$, $D_i$ its degradation rate and $Q_{ij}$ the probability that an individual of type $I_j$ mutates into type $I_i$ as a result of an unexact replication

$$\sum_{i=1}^{K} Q_{ij} = 1. \tag{1}$$

In this model selection happens at discrete time steps $h$. Between a selection event and the successive one, the organisms of the different types $I_j$ reproduce, mutate and degradate with their characteristic rates. We assume that only the individuals in the parental generation are subject to degradation, while the newborns will always reach the next selection step, which will restore the total population to $N$.

Let us denote with $\mathbf{n} = (n_1, \ldots, n_K)$, $\sum_{j=1}^{K} n_j = N$, the type counts just after a selection event, then, according to the above hypotheses, the expected number of individuals of type $I_i$ just before the next selection event will be given by

$$m_i = n_i + h \sum_{j=1}^{K} \left( A_j Q_{ij} - D_i \delta_{ij} \right) n_j. \tag{2}$$

All the quantities appearing in the above equation, with the exception of the $n_i$ that are integer numbers, can assume real values. Indeed, equation (2) should be interpreted as the deterministic limit of a stochastic process (see also [13]). Notice that, since in (2) the quantity $hD_i n_i$ represents the number of $I_i$ individuals in the parental generation that die before the next selection event, the time step length $h$ is bounded by the conditions $hD_i \leq 1$, $i = 1, \ldots, K$.

Selection consists in the extraction with replacement of $N$ individuals from the $\mathbf{m}$ population with sampling probabilities $\psi_i(\mathbf{n})$ equal to their relative frequencies

$$\psi_i(\mathbf{n}) := \frac{m_i}{\sum_{j=1}^{K} m_j} = \frac{n_i + h \sum_{j=1}^{K} \left( A_j Q_{ij} - D_i \delta_{ij} \right) n_j}{N + h \sum_{j=1}^{K} \left( A_j - D_j \right) n_j}. \tag{3}$$

Notice that, by definition, $\sum_i \psi_i(\mathbf{n}) = 1$ and that $\psi_i(\mathbf{n}) \geq 0$ is granted by the conditions $hD_i \leq 1$, so that the interpretation of the $\psi_i(\mathbf{n})$ as probabilities is adequate.

The Markov matrix of the model will be given by

$$P_h(\mathbf{n}'|\mathbf{n}) = \frac{N!}{n_1'! \ldots n_K'!} \psi_1(\mathbf{n})^{n_1'} \ldots \psi_K(\mathbf{n})^{n_K'}, \tag{4}$$

where $\mathbf{n} = (n_1, \ldots, n_K)$ and $\mathbf{n}' = (n_1', \ldots, n_K')$ are the type counts in successive generations, with $\sum_{j=1}^K n_j = \sum_{j=1}^K n_j' = N$.

Through (3) and (4), we have defined a family of stochastic models parametrized by the time step $h$. The particular case when

$$hD_j = 1, \ \forall j, \tag{5}$$

corresponds to the case of separated generations, when all the individuals in the parental generation die before the next reproductive step. When (5) holds, the sampling probabilities (3) simplify to

$$\psi_i(\mathbf{n}) := \frac{\sum_{j=1}^K \left(A_j Q_{ij} n_j\right)}{\sum_{j=1}^K A_j n_j}. \tag{6}$$

In this case, equation (4) defines the asexual Wright-Fisher model (see, for example, [1]), with the viability of type $I_i$ given by $A_i$. Notice also that the model is now independent of the time step $h$. We conclude that the stochastic model defined by (4) reduces to the asexual haploid Wright-Fisher model when the generations are separated.

Let us now consider the deterministic limit of the model (4). First of all, let us notice that the probability that $n_i' = k$ is simply given by

$$P_h(n_i' = k|\mathbf{n}) = \binom{N}{k} \psi_i(\mathbf{n})^k \left(1 - \psi_i(\mathbf{n})^{N-k}\right). \tag{7}$$

Accordingly, the expected value of $n_i'$ will be given by

$$\bar{n}_i' = \sum_{k=1}^K k \binom{N}{k} \psi_i(\mathbf{n})^k \left(1 - \psi_i(\mathbf{n})^{N-k}\right). \tag{8}$$

Using the binomial identity

$$\sum_{k=1}^K k \binom{N}{k} x^k y^{N-k} = Nx(x+y)^{N-1}, \tag{9}$$

we get

$$\bar{n}_i' = N\psi_i(\mathbf{n}). \tag{10}$$

By substituting the expression for the sampling probability (3) with $\mathbf{n}$ replaced by its expected value $\bar{\mathbf{n}}$ inside (10), we get the following system of discrete equations

$$\bar{n}_i' = \frac{\bar{n}_i + h \sum_{j=1}^K \left(A_j Q_{ij} - D_i \delta_{ij}\right) \bar{n}_j}{1 + h/N \sum_{j=1}^K \left(A_j - D_j\right) \bar{n}_j}. \tag{11}$$

Dividing equation (11) by $N$ we obtain the equations for the type frequencies $\phi_i = \bar{n}_i/N$:

$$\phi_i' = \frac{\phi_i + h \sum_{j=1}^{K} \left(A_j Q_{ij} - D_i \delta_{ij}\right) \phi_j}{1 + h \sum_{j=1}^{K} \left(A_j - D_j\right) \phi_j}. \tag{12}$$

For $h \to 0$, we have the following asymptotic expansion

$$\phi_i' = \left[\phi_i + h \sum_{j=1}^{K} \left(A_j Q_{ij} - D_i \delta_{ij}\right) \phi_j\right] \left\{1 - h \left[\sum_{j=1}^{K} \left(A_j - D_j\right) \phi_j\right] + O(h^2)\right\}, \tag{13}$$

from which it follows

$$\frac{\phi_i' - \phi_i}{h} = \sum_{j=1}^{K} \left(A_j Q_{ij} - D_i \delta_{ij}\right) \phi_j - \phi_i \left[\sum_{j=1}^{K} \left(A_j - D_j\right) \phi_j\right] + O(h). \tag{14}$$

The Eigen model equations are obtained by taking the limit $h \to 0$ of the above expression

$$\frac{d\phi_i}{dh} = \sum_{j=1}^{K} \left(A_j Q_{ij} - D_i \delta_{ij}\right) \phi_j - \frac{\phi_i}{N} \left[\sum_{j=1}^{K} \left(A_j - D_j\right) \phi_j\right]. \tag{15}$$

Clearly, by imposing separated generations (5) into equation (11) we would obtain the deterministic limit of the haploid asexual Wright-Fisher model, that coincides with the classical haploid mutation-selection model (see, for example, [3]):

$$\phi_i' = \frac{\sum_{j=1}^{K} A_j Q_{ij} \phi_j}{\sum_{j=1}^{K} A_j \phi_j}. \tag{16}$$

So, the only differences between the classical haploid mutation-selection model (16) and the Eigen model (15) is that the first one is obtained by the deterministic model (11) imposing separated generations and the latter taking the continuous time limit. Since neither the error threshold, nor the quasispecies phenomenon are due to these differences, they are naturally present in both models (indeed, see [3] for the quasispecies and [16] for the error threshold in the context of population genetics).

Just to give a concrete example, let us compute the error threshold according to both models in a very special case that allows for a simple analytical treatment. Let us suppose that each type $I_j$ is specificated by a genotype of length $L$ (so that $K = 4^L$). Let $u$ be the point mutation probability and let us send $u \to 0$ and $L \to \infty$ in such a way that the genomic mutation rate $U = uL$ stays finite. In this limit, the probability of mutation from the type $I_1$ to a different type $I_j$, $j \neq 1$ will be given by $\mu = 1 - \exp(-U)$ and the probability of back mutation will be zero. Let us also consider the single

peak fitness landscape $A_i = A_2$, $D_i = D_2$, $i > 2$, $A_1 - D_1 > A_2 - D_2$. The single peak fitness landscape is a (very) simplified fitness landscape often used in the Eigen model to get an analytical expression for the error threshold (see [14]). Let $\phi_B$ be the frequency of all the sequences different from $\phi_1$:

$$\phi_B = \sum_{i=2}^{\infty} \phi_i, \tag{17}$$

then the Eigen equations reduce to only two equations:

$$\dot{\phi}_1 = \left( A_1 e^{-U} - D_1 \right) \phi_1 - \phi_1 \left[ \left( A_1 - D_1 \right) \phi_1 + \left( A_2 - D_2 \right) \phi_B \right], \tag{18}$$

$$\dot{\phi}_B = A_1 \left( 1 - e^{-U} \right) \phi_1 + (A_2 - D_2)\phi_M - \phi_B \left[ \left( A_1 - D_1 \right) \phi_1 + \left( A_2 - D_2 \right) \phi_B \right]. \tag{19}$$

The error threshold corresponds to the smallest value of the genomic mutation rate $U$ for which

$$\lim_{t \to \infty} \phi_1(t) = 0. \tag{20}$$

From equations (18), (19), we get the error threshold

$$U_t = \ln \left( \frac{A_1}{A_2 - D_2 + D_1} \right). \tag{21}$$

For the classical haploid mutation-selection model (16) the above assumptions translate in considering a locus with two alleles of relative viability $w_1 = A_1 > A_2 = w_2$, with forward mutation given by $\mu = 1 - \exp(-U)$ and the probability of back mutation being zero. The frequencies $\phi_1'$ and $\phi_2'$ of the two alleles at the next generation, given that they are $\phi_1$ and $\phi_2$ at the present one will be:

$$\phi_1' = \frac{\phi_1 A_1 (1 - \mu)}{\phi_1 A_1 + \phi_2 A_2}, \tag{22}$$

$$\phi_2' = \frac{\phi_1 A_1 \mu + \phi_2 A_2}{\phi_1 A_1 + \phi_2 A_2}. \tag{23}$$

The first allele will go extinct, in the asymptotic limit, when $\phi_1' - \phi_1 < 0$ for $\phi_1 > 0$, that implies

$$\mu > \frac{A_1 - A_2}{A_1}, \tag{24}$$

or

$$U_t = \ln \left( \frac{A_1}{A_2} \right). \tag{25}$$

6

This is equivalent to the Eigen model result by keeping into account that the separated generations condition (5) implies that $D_1 = D_2$. By rescaling $A_1$ to 1 and $A_2$ to $1 - s$, where $s$ is the selection coefficient, we obtain:

$$U_t = \ln\left(\frac{1}{1-s}\right). \tag{26}$$

# 3 Error threshold in the sexual diploid Wright-Fisher model

Given the relation between the Eigen model (15) and the classical haploid mutation-selection model (16), it seems a natural option to use the classical diploid mutation-selection equation to determine the error threshold for sexual diploids. This was indeed done in [16], but the analytical derivation of the error threshold was inaccurate, giving the correct value of the critical mutation probability only for some regions of the $(h, s)$ space, where $h$ is the dominance and $s$ the selection parameter. To explain the problem with the derivation given in [16] let us briefly recall it. To obtain an analytic expression for the error threshold, the authors consider a diploid analogue of the simplifying assumptions that we used in the previous section. Namely, they considered a single locus with two alleles with mutation probability from the fittest to the worst allele given by $1 - m_{11}$ and vanishing back mutation probability (we recall that these simplifying assumptions comes from considering a genome of infinite length in the single peak landscape, see the previous section). Using these assumptions, the continuous time version of the classical diploid mutation-selection equation for the master frequency $x_1$, reduces to (compare with eq. (1) in [16]):

$$\dot{x}_1 = x_1(Wx)_1 m_{11} - x_1(x, Wx) \tag{27}$$

where (in their notations) $x = (x_1, 1 - x_1)$ is the vector of frequencies, $W$ the viabilities matrix and $(1 - m_{11})$ is the mutation probability of the master sequence. Next, they looked for a stationary solution of equation (27) (see eq. (17) in [16]):

$$x_1(Wx)_1 m_{11} - x_1(x, Wx) = 0, \tag{28}$$

When solving equation (28) for $m_{11}$, they neglected the common $x_1$ factor (see eq. (18) in [16]). That is, they solved the equation

$$(Wx)_1 m_{11} - (x, Wx) = 0 \tag{29}$$

for $m_{11}$ in the case $x_1 = 0$. However, this procedure is incomplete for two reasons. First, it should be checked that for the obtained value of $m_{11}$ there are no more solutions of equation (29) for $x_1 \in (0, 1]$, because in this case $x_1 = 0$ would not be a global sink for the equation (27). Second, if $m_{11}$ is such that the left hand side of equation (29) is not zero but always negative, $x_1 = 0$ will be a sink for equation (27) and the corresponding value of $m_{11}$ a candidate for the error threshold even if equation (29) is not satisfied for $x_1 = 0$. So, in the following we will determine the error threshold by keeping into account the above considerations.

To the effect of determining the error threshold, it is equivalent to consider the time discrete or continuous, so that we will use the more traditional discrete time version of the classical diploid mutation-selection equation. As usual, we will consider two alleles on an autosomal locus in a monoecius random mating population with separated generations. Let us denote by $A$ the fittest allele and by $p$ its frequency after the mutation step but before selection, while the frequency of the other allele $a$ will be given by $1 - p$. Let the relative fitness be given by 1 for $AA$, $1 - hs$ for $Aa$ and $1 - s$ for $aa$. We will denote by $\mu$ the mutation probability from $A$ to $a$ and, using the same approximation that we considered in the haploid case, we will set the back mutation probability to zero. Furthermore, we will restrict our considerations to the case when $0 < h < 1$, that is, we will neglect underdominance and overdominance. The frequency $p'$ of the $A$ allele after a generation (composed by selection followed by mutation) will be given by (see [3]):

$$p' = \frac{(1 - \mu) \left[ p^2 + p(1 - p)(1 - hs) \right]}{\bar{w}} \tag{30}$$

where $\bar{w}$ is the average fitness:

$$\bar{w} = p^2 + 2p(1 - p)(1 - hs) + (1 - p)^2(1 - s) \tag{31}$$

We want to determine the minimum value of the mutation probability $\mu$ that determines the extinction of the fittest allele $A$ from the population given that its initial frequency is $p_0 = 1$. So, we need to find the minimum value of $\mu$ that implies $p' < p$ for any $p$. Since $\bar{w} > 0$ this gives the inequality:

$$p(1 - \mu) \left[ p + (1 - p)(1 - hs) \right] - \bar{w}p < 0 \tag{32}$$

8

By eliminating the common $p$ factor we reduce to the quadratic inequality in $p$:

$$ap^2 + bp + c < 0 \qquad p \in (0, 1], \tag{33}$$

with

$$a \; = \; s(1 - 2h), \tag{34}$$

$$b \; = \; hs(1 - \mu) - 2s(1 - h), \tag{35}$$

$$c \; = \; (1 - hs)(1 - \mu) - (1 - s). \tag{36}$$

For $h < 1/2$, both the $a$ coefficient (34) and the discriminant $\Delta = b^2 - 4ac$ are positive. In this case, the parabola $ap^2 + bp + c$ will always have two real roots and will be negative in the region among the roots. One of the roots will be zero when $c = 0$, that is when

$$\mu = \frac{(1 - h)s}{1 - hs}. \tag{37}$$

For this value of $\mu$ the second root becomes

$$\frac{3h - 2 + hs(1 - 2h)}{(1 - hs)(2h - 1)}. \tag{38}$$

This last quantity will be greater than 1 for $h < 1$, that is always satisfied. We conclude that, for $h < 1/2$, the error threshold is given by equation (37). In the case $h = 1/2$ the equation (37) remains valid by continuity. Alternatively, since $a = 0$ one can directly solve the inequality

$$bp + c < 0, \tag{39}$$

that implies

$$u > \frac{1/2s}{1 - 1/2s}. \tag{40}$$

When $h > 1/2$, $a < 0$ and the discriminant can be both positive or negative. The inequality (33) will be satisfied if one of these three conditions is satisfied:

1. $\Delta \geq 0$ and the largest root is less than or equal to zero,

2. $\Delta \geq 0$ and the smallest root is greater than or equal to one,

9

3. $\Delta < 0$.

Equation (37) implies that one of the roots is zero, so that the first condition is satisfied if the second root (38) is less than or equal to zero. This is the case if

$$s > \frac{3h - 2}{h(2h - 1)}. \tag{41}$$

Notice that the right hand side of (41) is less than zero when $1/2 < h < 2/3$ and it is greater than zero but less than one when $2/3 < h < 1$.

The second condition can never be satisfied. Indeed, a necessary condition for the smallest root of a parabola to be greater than one is that also the abscissa value of the vertex $x = -b/(2a)$ be greater than one. In our case this translates into the condition

$$-hs(u + 1) > 0, \tag{42}$$

that cannot be satisfied for our choice of the range of the parameters.

Regarding the third condition, there exists real $\mu$ solutions to $\Delta < 0$ only when

$$s < \frac{2h - 1}{h^2}. \tag{43}$$

Since it holds

$$\frac{2h - 1}{h^2} \geq \frac{3h - 2}{h(2h - 1)}, \qquad \frac{1}{2} < h < 1, \tag{44}$$

the two regions (41) and (43) cover all the region $0 < s < 1$, $1/2 < h < 1$. Solving $\Delta < 0$ and imposing $\mu < 1$, we get the solution:

$$\mu > \frac{2(2h - 1) - h^2 s - 2\sqrt{(1 - 2h)(1 - 2h + h^2 s)}}{h^2 s}, \qquad \frac{1}{2} < h < 1, \quad 0 < s < \frac{2h - 1}{h^2}. \tag{45}$$

In the region

$$\max\left(0, \frac{3h - 2}{h(2h - 1)}\right) < s < \frac{2h - 1}{h^2} \qquad \frac{1}{2} < h < 1 \tag{46}$$

we have the two possible solutions for the error threshold:

$$\mu_1 = \frac{(1 - h)s}{1 - hs} \tag{47}$$

$$\mu_2 = \frac{2(2h - 1) - h^2 s - 2\sqrt{(1 - 2h)(1 - 2h + h^2 s)}}{h^2 s} \tag{48}$$

10

The two solutions $\mu_1$ and $\mu_2$ have the same value on the curve

$$\mu_1 = \mu_2 \quad \text{for} \quad s = \frac{3h-2}{h(2h-1)}. \tag{49}$$

Notice that the curve for $s$ (49) assumes positive values only when $h > 2/3$. Since a continuous solution for the error threshold must exist in the entire region $0 < h < 1$, $0 < s < 1$, we conclude that the error threshold will be given by:

$$\begin{cases} \mu_{\text{sex}} = \mu_1 & 0 < h \le \frac{2}{3}, \ 0 < s < 1 \text{ or } \frac{2}{3} < h < 1, \ \max\left(0, \frac{3h-2}{h(2h-1)}\right) < s < 1 \\ \mu_{\text{sex}} = \mu_2 & \frac{2}{3} < h < 1, \ s < \frac{3h-2}{h(2h-1)} \end{cases} \tag{50}$$

The result given in [16] coincides with the first line of equation (50).

# 4    The asexual diploids case

It is interesting to compare the result for the error threshold of sexual diploids (50) obtained in the previous section with that of asexual diploids, to evaluate the effect of syngamy on the error threshold. To this aim, we now calculate the error threshold for an asexual diploid organism, in the usual approximation of infinite genome length and single peak fitness landscape. If we have a diploid locus in an asexual organism, we denote with $p_1$, $p_2$ and $p_3$ respectively the frequencies of the $AA$, $Aa$ and $aa$ genotypes and with $\mu$ the probability of mutating from $A$ to $a$, then under the above hypotheses, we have that after one generation:

$$\vec{p}\,' = M\vec{p} \tag{51}$$

with $M$ given by

$$\begin{pmatrix} (1-\mu)^2 & 0 & 0 \\ 2\mu(1-\mu)(1-hs) & (1-\mu)(1-hs) & 0 \\ \mu^2(1-s) & \mu(1-s) & 1-s \end{pmatrix} \tag{52}$$

Since the equations are linear, there is no need of normalizing. The asymptotic frequencies will be given by $p_i^{(\infty)}/(p_1^{(\infty)} + p_2^{(\infty)} + p_3^{(\infty)})$. The outcome will depend on which is the maximum eigenvalue of the matrix $M$. If the maximum eigenvalue is $(1-\mu)^2$, then the three genotypes will coexist, because the corresponding eigenvector of $M$ has its three components different from zero. If the maximum

eigenvalue is $(1 - \mu)(1 - hs)$, then the $AA$ omozygote will disappear and the other two will coexist, because the corresponding eigenvector of $M$ has the first component zero and the other two different from zero. Finally, if the maximum eigenvalue is $1 - s$, then only the homozygote $aa$ will survive, because the corresponding eigenvector of $M$ has only its third component different from zero. The first case occurs when

$$\mu < \min\left(hs, 1 - \sqrt{1-s}\right). \tag{53}$$

The second case when

$$\frac{(1-h)s}{1-hs} > \mu > hs. \tag{54}$$

Notice that the condition

$$\frac{(1-h)s}{1-hs} > hs \tag{55}$$

implies

$$s > \frac{2h-1}{h^2}. \tag{56}$$

Accordingly, the threshold mutation rate for the loss of the homozygote $AA$ is given by

$$\mu_{p_1=0} = hs \qquad \max\left(0, \frac{2h-1}{h^2}\right) < s < 1. \tag{57}$$

Finally when

$$\mu > \max\left(\frac{(1-h)s}{1-hs}, 1 - \sqrt{1-s}\right). \tag{58}$$

there is the complete loss of the advantageous allele. Notice that

$$\frac{(1-h)s}{1-hs} > 1 - \sqrt{1-s} \quad \Leftrightarrow \quad h < \frac{1-\sqrt{1-s}}{s}, \tag{59}$$

and that

$$\frac{1}{2} \leq \frac{1-\sqrt{1-s}}{s} \leq 1, \quad 0 \leq s \leq 1. \tag{60}$$

So, the error threshold will be given by

$$\mu_{p_1=p_2=0} = \begin{cases} \frac{(1-h)s}{1-hs} & h < \frac{1-\sqrt{1-s}}{s}, \\ 1 - \sqrt{1-s} & h \geq \frac{1-\sqrt{1-s}}{s}. \end{cases} \tag{61}$$

We can now compare the error thresholds in the sexual and asexual case. Comparing equations (50), (61) and keeping into account (59) and (60), we see that for $h \leq 1/2$ there is no difference in the error threshold between the sexual and asexual case. We plot this difference in figure 1 for the whole range of variation of $h$ and $s$. We see that, for $h > 1/2$, the advantageous allele is more robust to complete loss by mutation in the asexual than in the sexual case.

In figure 2 we show the difference between the sexual error threshold (50) and the threshold mutation rate for the loss of the advantageous homozygote in the asexual case. We see that in this case, as obvious, the advantageous homozygote is much more robust to loss by mutation in the sexual than in the asexual case. Indeed, in the sexual case the advantageous homozygote can be eliminated only by completely removing the advantageous allele.

# 5    Conclusions

We constructed a stochastic model having the haploid Wright-Fisher model and the Eigen model as particular subcases. The haploid Wright-Fisher model is obtained by considering separated generations, while the Eigen model is obtained by taking the deterministic and continuous time limit. This derivation makes it clear what are the differences between these two important models of mutation-selection dynamics. Emerging as a deterministic limit, the Eigen model neglects genetic drift and it is almost equivalent to the deterministic limit of the haploid Wright-Fisher model, that is, the classical haploid mutation-selection model (16). The differences among this model and the Eigen model do not invalidate the concepts of quasispecies and error threshold, that consequently are present in both models. This suggests to use the classical diploid mutation-selection model to obtain the error threshold for sexual diploids. We derived an analytical expression for the error threshold inside this model by using the usual approximations of infinite genome length and the single peak fitness landscape. We compared this expression with the corresponding expression for asexual diploid organisms. No difference emerges when $h \leq 1/2$, but, curiously, when $h > 1/2$, syngamy makes the advantageous allele more liable to complete loss by mutation. On the other hand, this is not the case for the loss of the advantageous homozygote that in the sexual case, especially for low values of the dominance parameter $h$ and high values of the selection coefficient as can be appreciated in figure

# References

[1] Baake, E., Gabriel, W., (2000) Biological evolution through mutation, selection, and drift: An introductory review. Ann. Rev. Comp. Phys. VII, 203–264.

[2] Bull, J.J., Meyers, L.A., Lachmann M., (2005) Quasispecies made simple. PLoS Comput. Biol. 1:e61.

[3] Burger, R., (1998) Mathematical properties of mutation-selection models. Genetica 102/103, 279–298.

[4] Eigen, M., (1971), Selforganization of Matter and the Evolution of Biological Macromolecules. Naturwissenschaften 58, 465–523.

[5] Eigen, M., Schuster, P., (1977), The Hypercycle. A Principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. Naturwissenschaften 64, 541–565.

[6] Holland, J.J., (2006) Transitions in Understanding of RNA Viruses: A Historical Perspective. Curr. Top. Microbiol. Immunol. 299, 371–401.

[7] Holmes, E.C., Moya, A., (2002) Is the Quasispecies Concept Relevant to RNA Viruses? J. Virol. 76, 460–462.

[8] Holmes, E.C., (2010) The RNA Virus Quasispecies: Fact or Fiction? J. Mol. Biol. 400, 271-273.

[9] Jenkins, G.M., Worobey, M., Woelk, C.H., Holmes, E.C., (2001) Evidence for the non-quasispecies evolution of RNA viruses. Mol. Biol. Evol. 18, 987–994.

[10] Lauring, A.S., Andino, R., (2010) Quasispecies Theory and the Behavior of RNA Viruses. PLoS Pathog. 6, e1001005.

[11] Más, A., López-Galíndez, C., Cacho, I., Gómez, J., Martínez, M.A., (2010), Unfinished Stories on Viral Quasispecies and Darwinian Views of Evolution. J. Mol. Biol. 397, 865–877.

[12] Moya, A., Elena, S.F., Bracho, A., Miralles, R., Barrio, E., (2000) The evolution of RNA viruses: A population genetics view. Proc. Natl. Acad. Sci. USA 97, 6967–6973.

[13] Musso, F., (2011), A stochastic version of the Eigen model. Bull. Math. Biol. 73, 151–180.

[14] Nowak, M.A., (2006) Evolutionary Dynamics: Exploring the Equations of Life. Harvard University Press.

[15] Takeuchi, N., Hogeweg, P., (2007) Error-threshold exists in fitness landscapes with lethal mutants. BMC Evol. Biol. 7:15.

[16] Wiehe, T., Baake, E., Schuster, P., (1995) Error Propagation in Reproduction of Diploid Organisms A Case Study on Single Peaked Landscapes. J. theor. Biol. 177, 1–15.

[17] Wilke, C.O., (2005) Quasispecies theory in the context of population genetics. BMC Evol. Biol. 5:44.
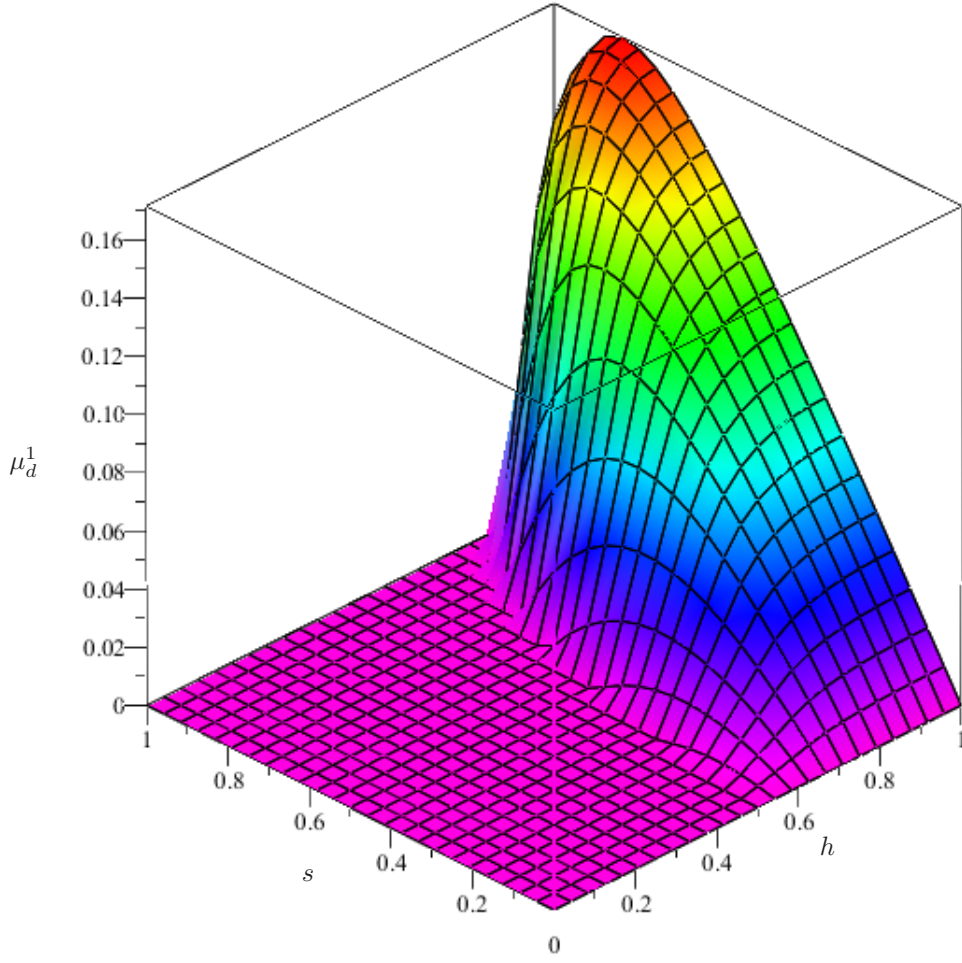
Figure 1: The difference $\mu_d^1 = \mu_{p_1=p_2=0} - \mu_{\text{sex}}$ between the mutation threshold for the complete loss of the advantageous allele in the asexual ($\mu_{p_1=p_2=0}$, see eq. (61)) and in the sexual case ($\mu$, see eq. (50)), versus the dominance $h$ and the selection coefficient $s$.
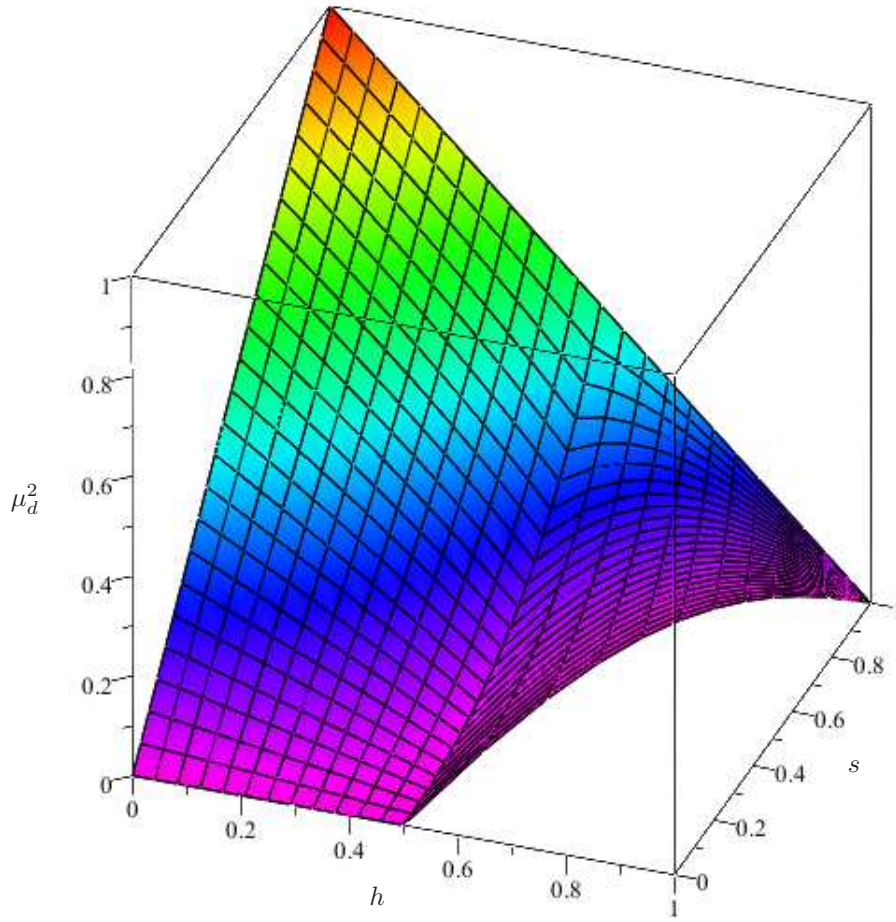
Figure 2: The difference $\mu_d^2 = \mu_{\text{sex}} - \mu_{p_1=0}$ between the mutation threshold for the loss of the advantageous homozygote in the sexual ($\mu$, see eq. (50)) and in the asexual case ($\mu_{p_1=0}$, see eq. (61)), versus the dominance $h$ and the selection coefficient $s$.